

RingMo: A Remote Sensing Foundation Model With Masked Image Modeling

Year: 2023 | Citations: 303 | Authors: Xian Sun, Peijin Wang, Wanxuan Lu, Zicong Zhu, Xiaonan Lu

Abstract

Deep learning approaches have contributed to the rapid development of remote sensing (RS) image interpretation. The most widely used training paradigm is to use ImageNet pretrained models to process RS data for specified tasks. However, there are issues such as domain gap between natural and RS scenes and the poor generalization capacity of RS models. It makes sense to develop a foundation model with general RS feature representation. Since a large amount of unlabeled data is available, the self-supervised method has more development significance than the fully supervised method in RS. However, most of the current self-supervised methods use contrastive learning, whose performance is sensitive to data augmentation, additional information, and selection of positive and negative pairs. In this article, we leverage the benefits of generative self-supervised learning (SSL) for RS images and propose an RS foundation model framework called RingMo, which consists of two parts. First, a large-scale dataset is constructed by collecting two million RS images from satellite and aerial platforms, covering multiple scenes and objects around the world. Second, we propose an RS foundation model training method designed for dense and small objects in complicated RS scenes. We show that the foundation model trained on our dataset with RingMo method achieves state-of-the-art (SOTA) on eight datasets across four downstream tasks, demonstrating the effectiveness of the proposed framework. Through in-depth exploration, we believe it is time for RS researchers to embrace generative SSL and leverage its general representation capabilities to speed up the development of RS applications.